Создание терминологического ресурса по катализу и его применение в системе текстового поиска

Альперин Борис Львович
Институт катализа им. Г.К. Борескова СО РАН (Новосибирск), Россия
e-mail: boris.alperin@gmail.com

Кузьмин Андрей Олегович
Институт катализа им. Г.К. Борескова СО РАН (Новосибирск), Россия

Эффективность научно-исследовательских работ напрямую зависит от качества их информационного обеспечения, а ключевыми этапами любого научного исследования являются поиск и представление информации, основой которых является использование и развитие терминологической базы предметной области.
Таким образом, создание и организация доступа к расширенным тезаурусам (глоссариям) в сети Интернет является актуальной задачей естественных наук.
В работе описана реализация терминологического веб-ресурса по катализу, доступного по адресу http://catalysis.ru/thezaurus/ и содержащего пользовательско-административную часть, а также систему веб-сервисов для работы с ним. Содержимое ресурса насчитывает около 20 тысяч понятий.
Интернет является на сегодняшний день важнейшим источником информации, однако поиск в нем специализированной научно-технической информации зачастую оказывается малоэффективным.
В работе разработаны методы использования терминологического ресурса для улучшения качества информационного поиска в Интернете путём использования тезауруса предметной области для уточнения поискового запроса пользователя.
При таком подходе в поисковом запросе выделяются термины предметной области, и далее запрос модифицируется с использованием связанных терминов из тезауруса так, чтобы они одновременно расширяли и сужали запрос, создавая альтернативу и исключая ненужные термины.
В работе рассматривается применение данного подхода с использованием тезауруса по катализу. Разработана методика создания расширенного поискового запроса на языке запросов Яндекс, с использованием различных типов связей между терминами тезауруса.
К настоящему времени разработано демонстрационное веб-приложение (http://catalysis.ru/thezaurus/queryimprover/), реализующее указанную методику модификации запросов.